DiG 10 (2002), 59-76

A Report on the *Corpus Oral Dialectal del Català Actual* (COD)

Maria-Rosa Lloret and *M. Pilar Perea*

1. Introduction

Since 1991 the Department of Catalan Philology has been preparing a set of linguistic corpora of contemporary Catalan. It is called *Corpus de la Universitat de Barcelona* (CUB), or the University of Barcelona Corpus. The aim of the project is to gather representative data in order to contribute to the study of language variation from a geographical, social, and functional perspective.*

The project is divided in two major areas of study: one is concerned with written data and the other with spoken data. The first one is called the Written Catalan Corpus (or CECA). It is based on information from newspapers. The second is called the Spoken Catalan Corpus (or COCA). It includes three groups dealing with functional, social, and geographical variation. Within the functional area of research (the COF), there are three subcorpora: the COM (or Media Corpus), which deals with data from radio and TV advertisements, the COC (or Speech Corpus), which studies textual phenomena and casual speech, and the COR (or Functional Varieties Corpus), which studies functional varieties and ethnographical aspects of speech acts. The two latter groups work together with the Social Varieties Corpus (the COS), and they form the group called CORCS (or Spoken Corpus of Functional Varieties, Speech, and Social Aspects). Finally, geographical variation is dealt with in the COD (or Dialectal Corpus), which provides data from the entire Catalan area. Figure 1 shows the general design of the University of Barcelona Corpus.

The general project was designed in three stages: data collection, data systematisation, and analysis of the results. In addition to that, a final stage of matching and comparing the results of the different subcorpora has also been planned.

The goal of this paper is to present a brief report on the Dialectal Corpus (COD). Concerning working procedures, and in line with the general project, the Dialectal Corpus involves three main steps: (i) collecting a spoken

* This project is sponsored by a DGICYT grant (PB97-0889) from the Spanish Government and by a CIRIT grant (SGR00411) from the Catalan Government.